



2022

人工智能课程 学习手册

科教兴国 | 科学强国



像素画 与 CV 算法

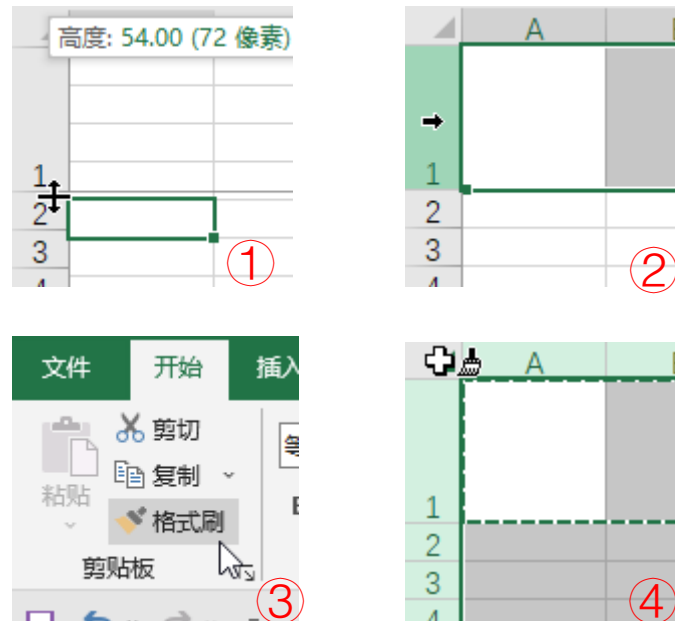
一、观察 Excel 的行高和列宽，为什么默认行高是 14.25，默认列宽是 8.38，而显然每一行的高度要小于每一列的宽度呢？

二、常规操作

1、将行高调整为 72 像素，使所有单元格为正方形，并通过格式刷布局整个表单

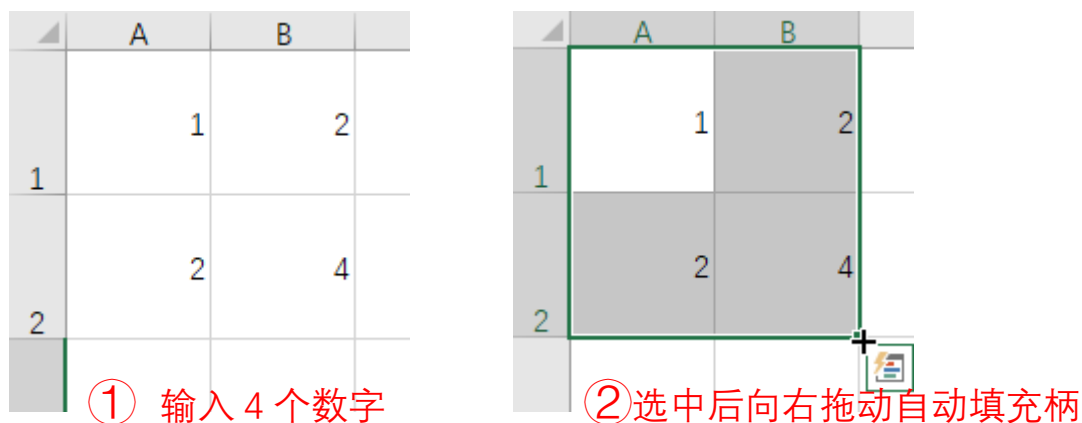
*可以通过拖动行间隔与列间隔调整行高和列宽

*可以通过单击格式刷将源格式复制到目的行/列/单元格中去



2、通过自动填充柄设置九九乘法表

*自动填充柄可以在行方向或列方向上进行序列填充，默认序列为等差数列



	A	B	C	D	E	F	G	H	I
1	1	2	3	4	5	6	7	8	9
2	2	4	6	8	10	12	14	16	18
3									
4									

③再向下拖动自动填充柄

3、使用条件格式填充数字区域

	A	B	C	D	E	F	G	H	I
1	1	2	3	4	5	6	7	8	9
2	2	4	6	8	10	12	14	16	18
3	3	6	9	12	15	18	21	24	27
4	4	8	12	16	20	24	28	32	36
5	5	10	15	20	25	30	35	40	45
6	6	12	18	24	30	36	42	48	54
7	7	14	21	28	35	42	49	56	63
8	8	16	24	32	40	48	56	64	72
9	9	18	27	36	45	54	63	72	81

条件格式 套用 表格格式

常规 差 好

计算 检查单元格 解释

突出显示单元格规则(H)

最前/最后规则(T)

数据条(D)

色阶(S)

图标集(I)

新建规则(N)...

清除规则(C)

管理规则(R)...

样式

O P

绿 - 黄 - 红色阶

为单元格区域添加颜色渐变。颜色指明每个单元格值在该区域内的位置。

4、对每一行值进行求和

*选中 J1 在编辑框中输入公式，注意公式用“=”符号开头

	A	B	C	D	E	F	G	H	I	J
1	1	2	3	4	5	6	7	8	9	+C1+D1 +E1+F1 +G1+H1 +I1

=A1+B1+C1+D1+E1+F1+G1+H1+I1

*拖动自动填充柄，填充到整列

	A	B	C	D	E	F	G	H	I	J
1	1	2	3	4	5	6	7	8	9	45
2	2	4	6	8	10	12	14	16	18	90
3	3	6	9	12	15	18	21	24	27	135
4	4	8	12	16	20	24	28	32	36	180
5	5	10	15	20	25	30	35	40	45	225
6	6	12	18	24	30	36	42	48	54	270
7	7	14	21	28	35	42	49	56	63	315
8	8	16	24	32	40	48	56	64	72	360
9	9	18	27	36	45	54	63	72	81	405

仔细观察这一列的公式变化，随着行号的增加，公式中的行号也在不断递增

1	1	2	3	4	5	6	7	8	9	=A1+B1+C1+D1+E1+F1+G1+H1+I1	I
2	2	4	6	8	10	12	14	16	18	=A2+B2+C2+D2+E2+F2+G2+H2+I2	
3	3	6	9	12	15	18	21	24	27	=A3+B3+C3+D3+E3+F3+G3+H3+I3	

思考：如果在一行中横向拖动自动填充柄，公式会如何变化？

5、对每一列值进行求和

*选中 A10 单元格，单击求和函数并确定

	A	B	C	D	E	F	G	H	I	J
1	1	2	3	4	5	6	7	8	9	45
2	2	4	6	8	10	12	14	16	18	90
3	3	6	9	12	15	18	21	24	27	135
4	4	8	12	16	20	24	28	32	36	180
5	5	10	15	20	25	30	35	40	45	225
6	6	12	18	24	30	36	42	48	54	270
7	7	14	21	28	35	42	49	56	63	315
8	8	16	24	32	40	48	56	64	72	360
9	9	18	27	36	45	54	63	72	81	405
10	=SUM(A1:A9)									



*拖动自动填充柄，填充到整行

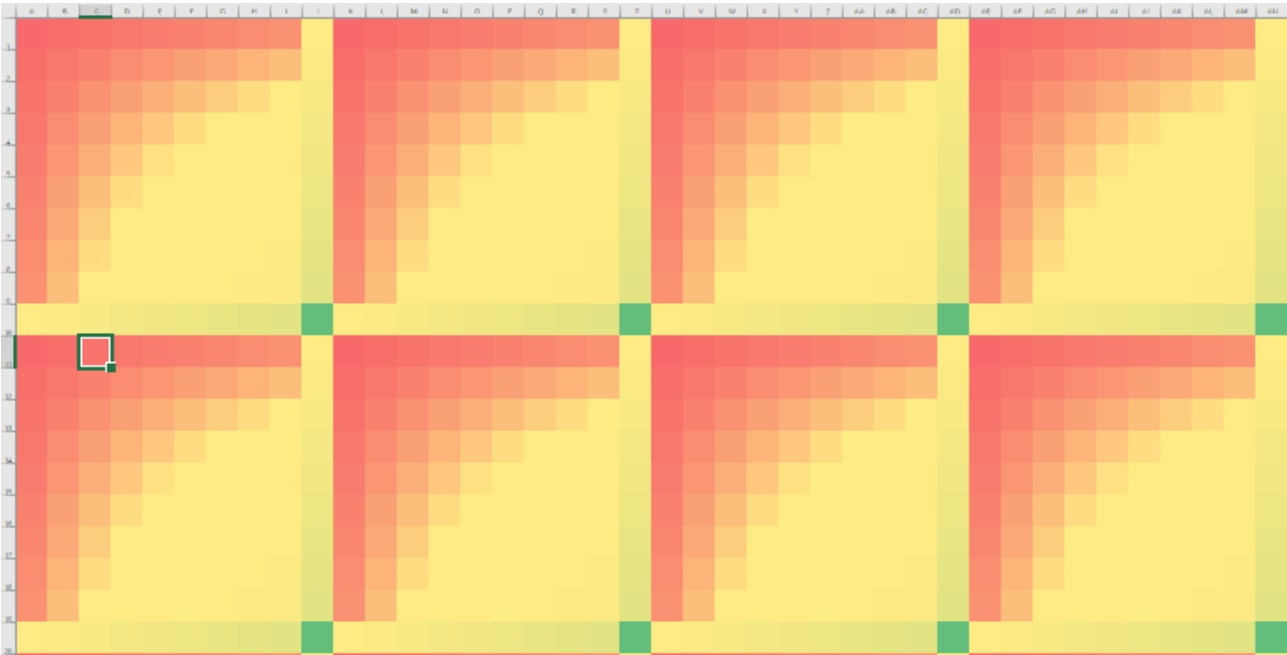
	A	B	C	D	E	F	G	H	I	J
1	1	2	3	4	5	6	7	8	9	45
2	2	4	6	8	10	12	14	16	18	90
3	3	6	9	12	15	18	21	24	27	135
4	4	8	12	16	20	24	28	32	36	180
5	5	10	15	20	25	30	35	40	45	225
6	6	12	18	24	30	36	42	48	54	270
7	7	14	21	28	35	42	49	56	63	315
8	8	16	24	32	40	48	56	64	72	360
9	9	18	27	36	45	54	63	72	81	405
10	45	90	135	180	225	270	315	360	405	2025

思考：

经过条件格式处理的单元格，如何将格式保留下来？

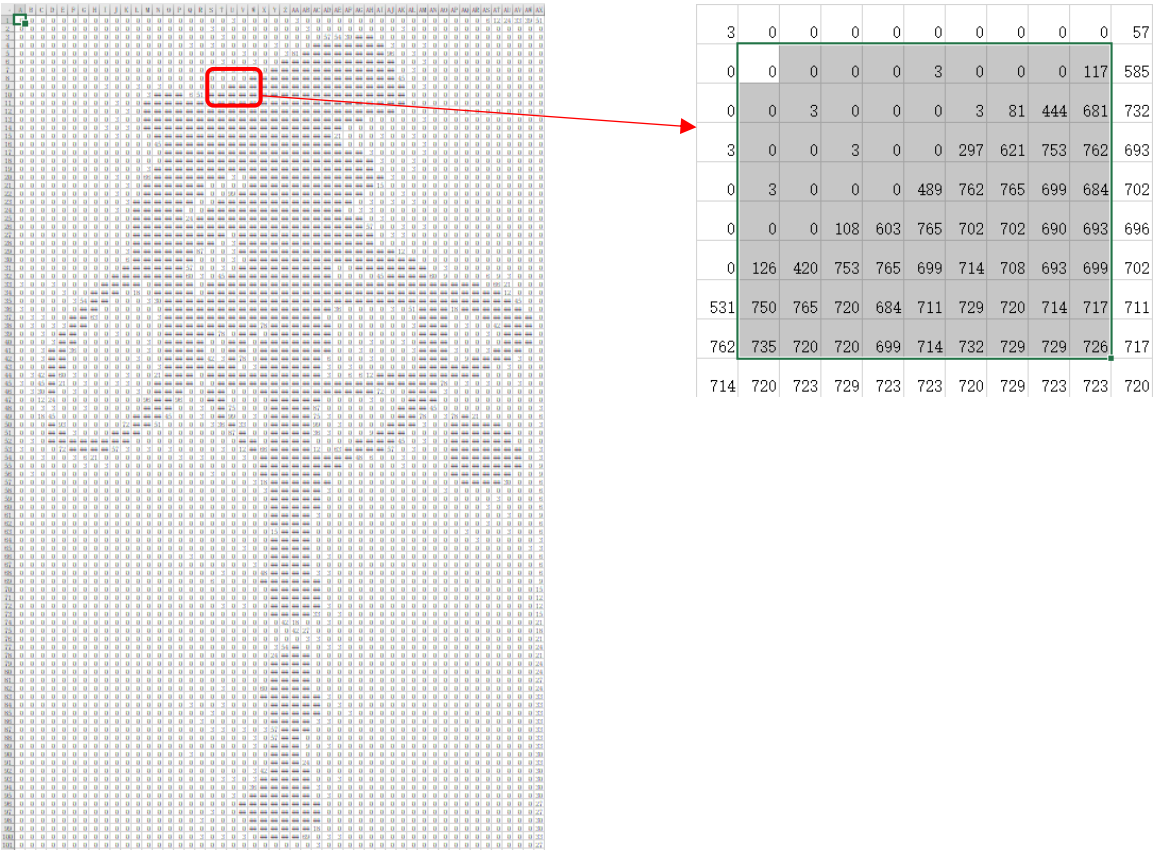
写出你的解决方案：

效果图如下：



三、数字画

在了解了通过不同的数字的条件格式设置可以在 Excel 中完成渐变色的填充，试着利用数字来绘制一幅数字画。



*参考



明 · 徐渭



明 · 祝允明



元 · 赵孟頫

四、像素画

运用格式刷绘制更为细腻的像素画

五、登录网站，尝试转换图像为 Excel 文件

什么是像素?

像素是指由图像的小方格组成的，这些小方格都有一个明确的位置和被分配的色彩数值，小方格颜色和位置就决定该图像所呈现出来的样子。可以将像素视为整个图像中不可分割的最小单位或者是元素。不可分割的意思是它不能够再切割成更小单位抑或是元素，它是以一个单一颜色的小格存在。每一个点阵图像包含了一定量的像素，这些像素决定图像在屏幕上所呈现的大小。

像素风格画

文件上传

上传图片

选择文件
 未选择任何文件
只限上传jpg文档，文件请用 英文命名，文件不超过1M.

下载默认文档

下载上传的文档

五、公式与函数

将九九乘法表中每一个值加上 20 以后，制作一张新表，如图所示：

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	1	2	3	4	5	6	7	8	9					21	22	23	24	25	26	27	28	29
2	2	4	6	8	10	12	14	16	18					22	24	26	28	30	32	34	36	38
3	3	6	9	12	15	18	21	24	27					23	26	29	32	35	38	41	44	47
4	4	8	12	16	20	24	28	32	36					24	28	32	36	40	44	48	52	56
5	5	10	15	20	25	30	35	40	45				+	20								
6	6	12	18	24	30	36	42	48	54					26	32	38	44	50	56	62	68	74
7	7	14	21	28	35	42	49	56	63					27	34	41	48	55	62	69	76	83
8	8	16	24	32	40	48	56	64	72					28	36	44	52	60	68	76	84	92
9	9	18	27	36	45	54	63	72	81					29	38	47	56	65	74	83	92	101

复习公式复制的方式：

$$N1 = A1 + L5$$

+1 ↓ ↓ +0

$$O1 = B1 + M5$$

新知识学习：

在复制单元格中公式的时候，可以用“\$”符号限制行号或列号的递增。

例如：

$$N1 = A1 + \$L5$$

+1 ↓ ↓ +0

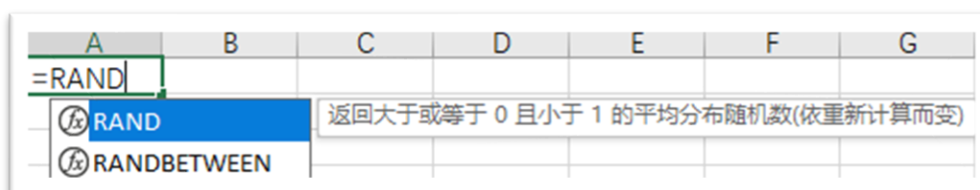
$$O1 = B1 + \$L5$$

思考：如果将 N1 的公式复制到 N2，仍然保持公式中 L5 不变，应该在哪里加上“\$”符号？复制完以后的公式是什么？

$$N1 = A1 + L5$$

$$N2 =$$

在 Excel 中，RAND 函数 返回了一个大于等于 0 且小于 1 的平均分布的随机实数。每次计算工作表时都会返回一个新的随机实数。



RANDBETWEEN(bottom, top) 函数返回位于两个指定数之间的一个随机整数。每次计算工作表时都将返回一个新的随机整数。



使用 RANDBETWEEN 函数，制作一张 9×9 的随机整数表（范围在 1~100 以内）

	A	B	C	D	E	F	G	H	I
1	26	38	94	10	87	96	87	80	77
2	34	37	85	71	11	87	65	74	9
3	68	63	42	10	80	100	86	20	56
4	51	65	46	9	36	68	73	80	28
5	32	26	56	88	21	10	87	95	74
6	6	65	29	91	36	21	86	26	11
7	83	7	33	46	48	3	4	20	83
8	6	88	66	24	31	42	5	54	38
9	14	88	41	24	85	35	59	58	65

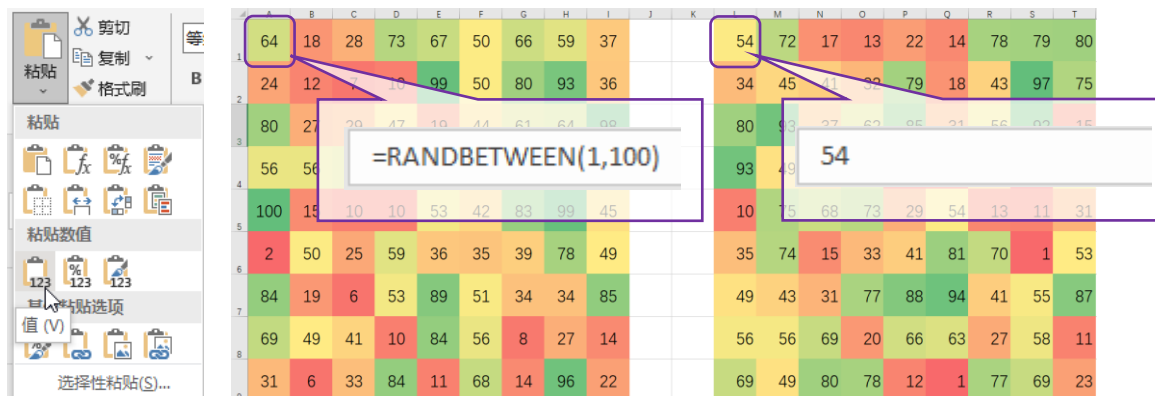
六、排序与筛选

选中 A1 单元格，点击排序和筛选，选中升序，查看排序以后的结果。

思考，为什么表格仍是随机排布的？



使用粘贴中的“值粘贴”方式，将函数值固定到新的表格中。



选中数据区域/单元格，点击排序和筛选，选中升序，查看排序以后的结果。（如下左图）

10	75	68	73	29	54	13	11	31
34	45	41	32	79	18	43	97	75
35	74	15	33	41	81	70	1	53
49	43	31	77	88	94	41	55	87
54	72	17	13	22	14	78	79	80
56	56	69	20	66	63	27	58	11
69	49	80	78	12	1	77	69	23
80	93	37	62	85	31	56	92	15
93	49	43	67	6	9	43	3	55

左图

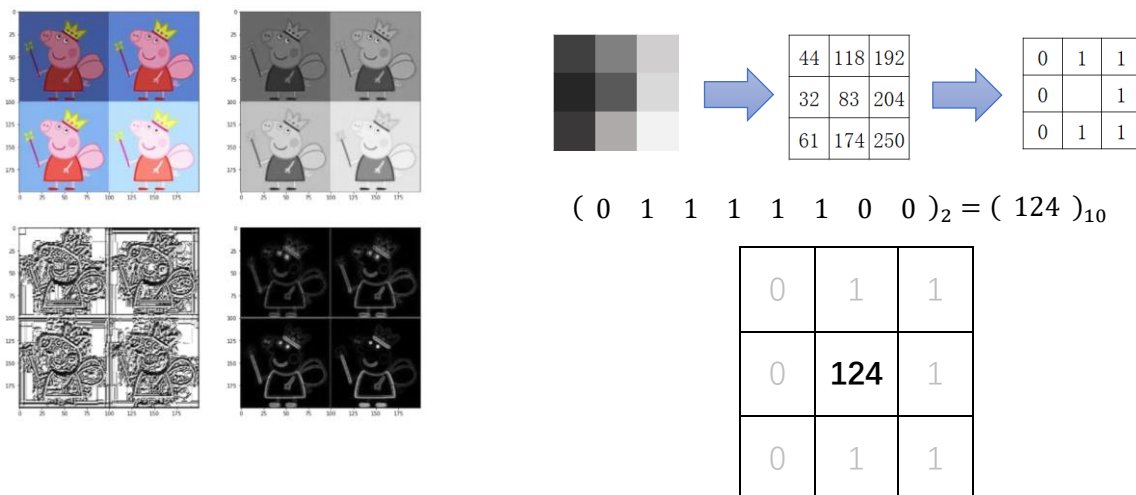
10	43	15	13	6	1	13	1	11
34	45	17	20	12	9	27	3	15
35	49	31	32	22	14	41	11	23
49	49	37	33	29	18	43	55	31
54	56	41	62	41	31	43	58	53
56	72	43	67	66	54	56	69	55
69	74	68	73	79	63	70	79	75
80	75	69	77	85	81	77	92	80
93	93	80	78	88	94	78	97	87

右图

思考：若要得到上右图的结果，应该如何调整排序范围？

七、LBP 算法

复习：人脸特征提取之 LBP 算法



LBP 算子定义为在 3×3 的窗口内，以窗口中心像素为阈值，将相邻的 8 个像素的灰度值与其进行比较，若周围像素值大于等于中心像素值，则该像素点的位置被标记为 1，否则为 0。这样， 3×3 邻域内的 8 个点经比较可产生 8 位二进制数（通常转换为十进制数即 LBP 码，共 256 种），即得到该窗口中心像素点的 LBP 值，并用这个值来反映该区域的纹理信息。需要注意的是，LBP 值是按照顺时针方向组成的二进制数。

试一试：

1、在一个 9×9 的区域中尝试使用 LBP 算法生成一幅新的图片。请计算并填写右图 4 个空格

	A	B	C	D	E	F	G	H	I
1	0	0	0	0	0	0	0	0	0
2	0	0	0	1	1	1	0	0	0
3	0	0	1	1	1	1	1	0	0
4	0	1	1	1	1	1	1	1	0
5	0	1	1	1	1	1	1	1	0
6	0	0	1	1	1	1	1	0	0
7	0	0	0	1	1	1	0	0	0
8	0	0	0	0	0	0	0	0	0

	L	M	N	O	P	Q
2			225	224	240	0
3			128	0	32	112
4	195	128	0	0	0	32
5	135	2	0	0	0	8
6	0	7	2	0	8	28
7	0	0	15	14	30	0

请写出这四个单元格中的算式：

$$L2 = 0 \times 128 + 0 \times 64 + 0 \times 32 + 0 \times 16 + 0 \times 8 + 0 \times 4 + 0 \times 2 + 0 \times 1$$

$$M2 =$$

$$L3 =$$

$$M3 =$$

如果使用 Excel 来进行 LBP 计算，则这四个单元格的公式如何编写？

方法一：

$$L2 =$$

$$M2 =$$

$$L3 =$$

$$M3 =$$

提示：要比较两个数值的大小，可以用数学方法进行运算，例如有 a 和 b 两个不相等的整数用 $(a-b) \div (a-b)$

可以计算得到 -1 和 1 两个可能的答案。

方法二：

L2 =

M2 =

L3 =

M3 =

提示：IF 函数是 Excel 中最常用的函数之一，它可以对值和期待值进行逻辑比较。

因此 IF 语句可能有两个结果。第一个结果是比较结果为 True，第二个结果是比较结果为 False。

例如，=IF(C2="Yes",1,2)

表示 IF(C2 = Yes, 则返回 1, 否则返回 2)。

上机练习，对下载的 Excel 数字图片进行 LBP 运算：

LBP算法简介

LBP (Local Binary Pattern, 局部二值模式) 是一种用来描述图像局部纹理特征的算子；它具有旋转不变性和灰度不变性等显著的优点。它是首先由T. Ojala, M. Pietikainen 和 D. Harwood 在1994年提出，用于纹理特征提取。

LBP特征的描述

原始的LBP算子定义为在3*3的窗口内，以窗口中心像素为阈值，将相邻的8个像素的灰度值与其进行比较，若周围像素值大于中心像素值，则该像素点的位置被标记为1，否则为0。这样，3*3邻域内的8个点经比较可产生8位二进制数（通常转换为十进制数即LBP码，共256种），即得到该窗口中心像素点的LBP值，并用这个值来反映该区域的纹理信息。

Example

10	12	15
11	15	22
30	3	54

Threshold

1	0	0
0	1	1
1	0	1

Pattern

(00010010) = (154)

文件上传

上传图片

选择文件

未选择任何文件

只限上传jpg文档，文件请用 英文命名，文件不超过1M.

不同亮度下的文件

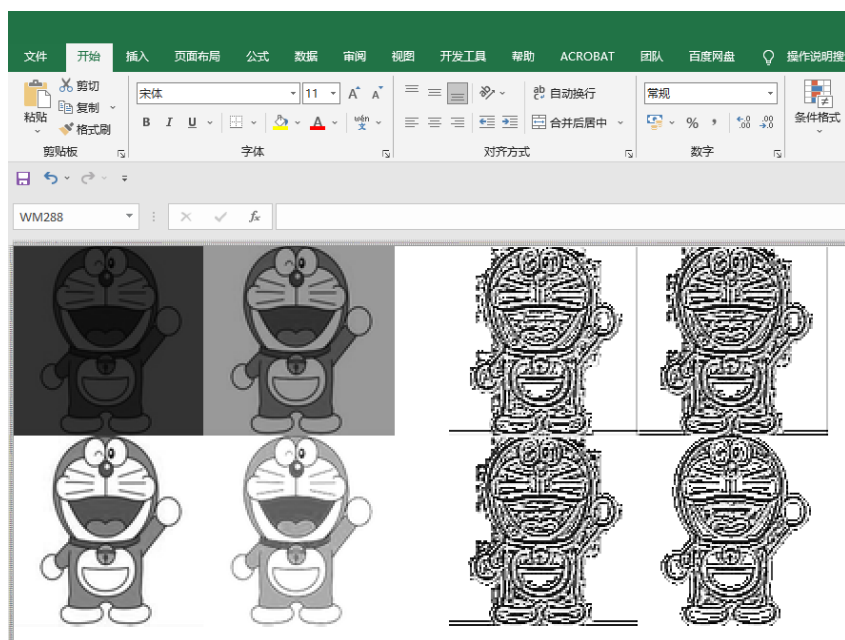






下载上传的文档





思考：

1、LBP 算法为什么能做到灰度不变性等优势特点？

2、为什么油画、照片用 LBP 处理以后效果不如漫画？

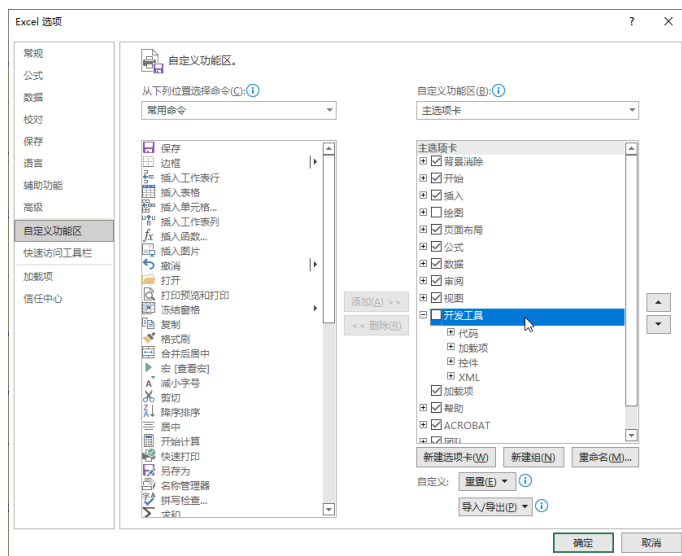
八、Excel 宏公式

VBA 是 Visual Basic for Application 的缩写。VBA 是在 Office 软件与 VB 的结合，是用于执行自动化任务的编程语言，VBA 适合处理复杂、重复性的操作。VBA 属于 VB 中的一种。

VBA 能够执行自动执行任务，将大量的重复过程用编程语言来描述，用户可以通过控件一键完成操作，从而大大地简化工作。



1、打开 VBA



打开文件→选项
选择“自定义功能区”
在右侧勾选“开发工具”



宏（英语：Macro）是一种批量处理的称谓。

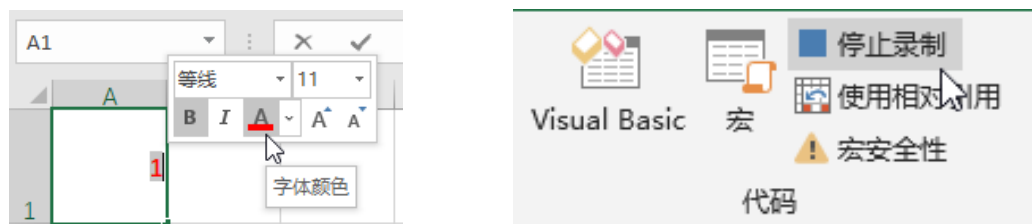
计算机科学里的宏是一种抽象（Abstraction），它根据一系列预定义的规则替换一定的文本模式。解释器或编译器在遇到宏时会自动进行这一模式替换。对于编译语言，宏展开在编译时发生，进行宏展开的工具常被称为宏展开器。

2、尝试“录制宏”

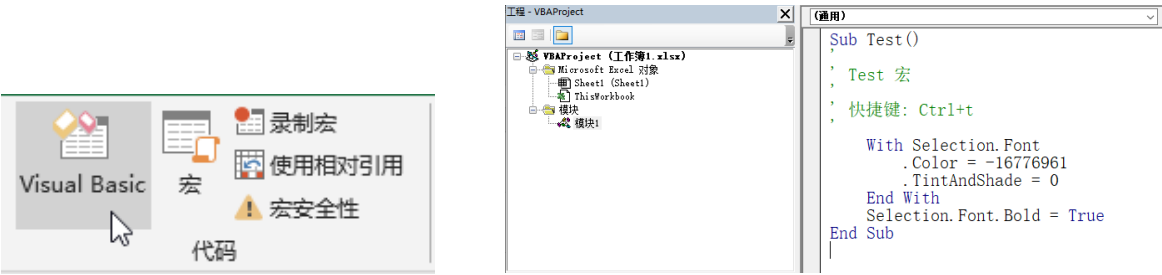
单击“录制宏” → 给宏定义名称“Test1” → 选择快捷键



设置 A1 单元格格式，然后单击“停止录制”



打开“Visual Basic”代码页面



使用设置的快捷键进行快速格式设置：

1	2	3	4	5	6	7	8	9
2	4	6	8	10	12	14	16	18
3	6	9	12	15	18	21	24	27
4	8	12	16	20	24	28	32	36
5	10	15	20	25	30	35	40	45
6	12	18	24	30	36	42	48	54
7	14	21	28	35	42	49	56	63
8	16	24	32	40	48	56	64	72
9	18	27	36	45	54	63	72	81

Ctrl+t

1	2	3	4	5	6	7	8	9
2	4	6	8	10	12	14	16	18
3	6	9	12	15	18	21	24	27
4	8	12	16	20	24	28	32	36
5	10	15	20	25	30	35	40	45
6	12	18	24	30	36	42	48	54
7	14	21	28	35	42	49	56	63
8	16	24	32	40	48	56	64	72
9	18	27	36	45	54	63	72	81

在开发工具下插入一个表单控件“按钮”，在表单上绘制出一个按钮，并指定执行宏为“Test1”



常用命令：

MsgBox ("finished")	弹出对话框
ActiveCell.Offset(i,j).Select	当前活动单元格移动 i 行 j 列并选中
Row = ActiveCell.Row	当前活动单元格的行号
Col = ActiveCell.Column	当前活动单元格的列号
ActiveCell.Value = 1	当前活动单元格的值设为 1
Cells(i,j).Value	设置/读取第 i 行 j 列的单元格数据
For i = 1 To 100	将语句 1 执行 100 次
语句 1	
Next	

思考：尝试用 VBA 制作一张乘法表？

八、卷积算法 (Convolution):

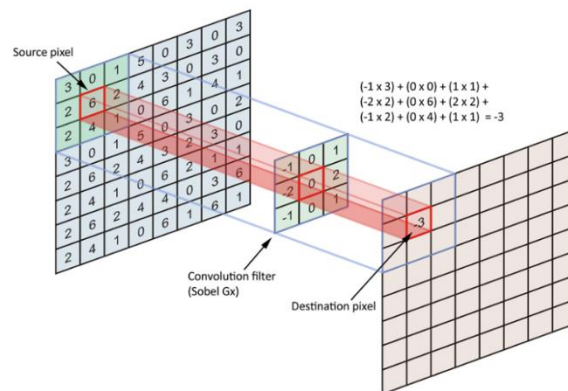
在 LBP 的运算中, LBP 算子以像素周围的值作为参考进行运算, 它的结果反映了该区域像素之间的关系。而正由于 LBP 算子来自图像本身, 所以 LBP 的运算结果也只能有一种, 即图像的纹理信息。如果有一个图像与来自图像外部的算子进行运算, 则运算结果会是什么样的呢?

新知识:

卷积和卷积核的概念:

用一个模板和一幅图像进行卷积, 对于图像上的一个点, 让模板的原点和该点重合, 然后模板上的点和图像上对应的点相乘, 然后各点的积相加, 就得到了该点的卷积值。

试将下左图 (原始图像) 通过卷积运算以后, 填入右图单元格中。猜猜看, 原始图像和卷积以后的图像分别是怎么样的。



0	0	0	0	0	0	0	0	0
0	0	0	1	1	1	0	0	0
0	0	1	1	1	1	1	0	0
0	1	1	1	1	1	1	1	0
0	1	1	1	1	1	1	1	0
0	0	1	1	1	1	1	0	0
0	0	0	1	1	1	0	0	0
0	0	0	0	0	0	0	0	0

-1	-1	-1
-1	8	-1
-1	-1	-1

卷积核

活动一:

运用 Excel 进行计算, 检验结果

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	0	0	0	0	0	0	0	0	0				
2	0	0	0	1	1	1	0	0	0				
3	0	0	1	1	1	1	1	0	0	-1	-1	-1	
4	0	1	1	1	1	1	1	1	0	-1	8	-1	
5	0	1	1	1	1	1	1	1	0	-1	-1	-1	
6	0	0	1	1	1	1	1	0	0				
7	0	0	0	1	1	1	0	0	0				
8	0	0	0	0	0	0	0	0	0				

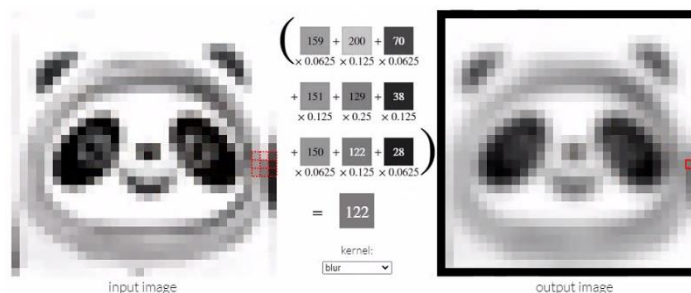
O	P	Q	R	S	T	U	V	W
=A1*\$K\$3+B1*\$L\$3+C1*\$M\$3+A2*\$K\$4+B2*\$L\$4+C2*\$M\$4+A3*\$K\$5+B3*\$L\$5+C3*\$M\$5								

活动二:

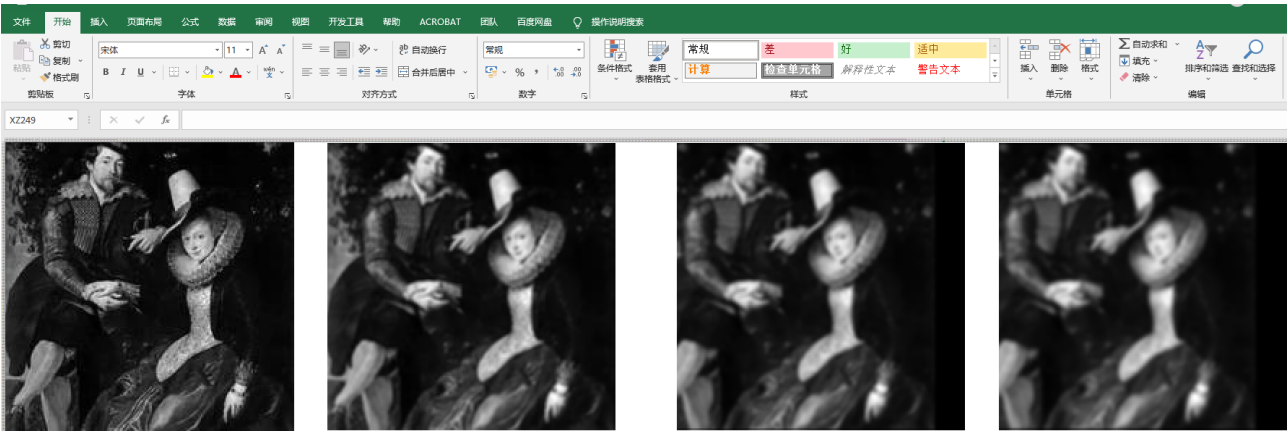
登录学习网站

<http://ai.sfls.cn> 的相应模块

进行卷积核设计



活动三：
下载数字画，使用卷积核进行运算



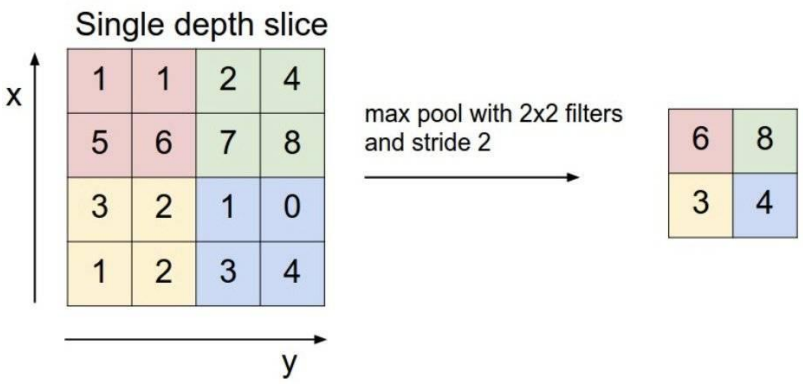
1.已知，每一幅像素画为 200 像素×200 像素
像素画左上角第一个单元格坐标为 A1，则像素画右下角最后一个单元格坐标为_____？
写出计算过程

如果在距离该像素画间隔 10 列建立卷积以后的新像素画，则经过卷积运算以后的像素画左上角第一个单元格的坐标为_____？
写出计算过程

假设有一个 3×3 的卷积核位于 GV1:GX3 单元格内，则在新像素画的第一个公式为

八、池化 (Pooling)

池化 (Pooling) 是卷积神经网络中的一个重要的概念，它实际上是一种形式的降采样。有多种不同形式的非线性池化函数，而其中“最大池化 (Max pooling)”是最为常见的。它是将输入的图像划分为若干个矩形区域，对每个子区域输出最大值。



项目活动：作业照片的优化处理

在线授课期间，许多教师布置了在线作业。同学们拍摄照片未经处理上传以后，给老师的批阅带来了很大的困难。在 app 市场上，有一些 App 可以快速处理图像。经过这一章的学习，我们学习了图像处理的一些方法。现在我们通过项目活动来将这些方法应用到实践中去。同时也研究一下各种图像处理软件的原理。

项目学习目标

本项目中，我们尝试利用 Excel 与在线平台，实现手写文字的处理。体验各种图像优化软件的处理过程与方法。
完成本项目学习以后，须能回答以下问题：

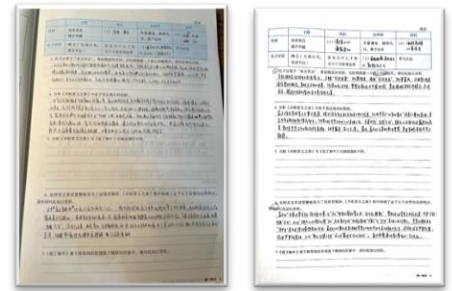
1. 如何正确摄制一幅图像？
2. 如何对图像进行二值化处理？
3. 如何运用卷积处理图像，不同的卷积核有什么作用？
4. 在 Excel 上如何进行池化操作，池化以后的图像有什么变化？

项目学习指引

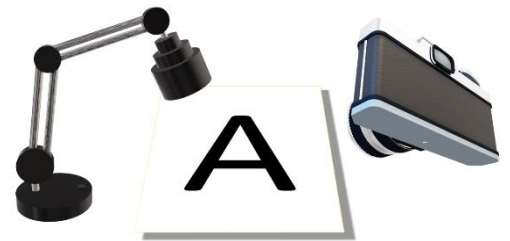
一、拍摄一张清晰的作业

拍摄作业的第一要务，是能让人看得清楚，感受准确。能看到整体与细节，所以照片要完整，干净，清晰，字迹不能变形。因此最重要的是距离和光线，拍摄一张完美的作业图片需要合理安排光线。其次是后期的调整和裁剪，将照片中多余的部分裁剪掉，这样作业才会更整洁。

在一张 A4 纸上写一个大写的字符，采用合适的光源，拍摄存储在计算机中。



两份不同的作业对比



二、转换图像为 Excel 数字格式

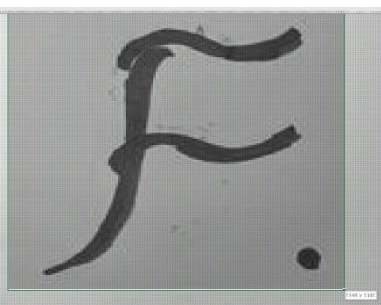
利用在线平台：<http://ai.sfls.cn> 相应模块

将图片上传以后转为 Excel 数字格式。（如果身边没有可用的图片，可以直接下载网站上默认素材）



三、对下载的 Excel 数字图像进行裁剪

1. 选中包含图像的正方形区域（例如 134 行×134 列）
2. 复制图像并粘贴到新工作表（完成裁剪操作）
3. 观察图像，是否需要旋转，如需要旋转则需选择转置粘贴



如果图像**无需**旋转

如果图像**需要**旋转

四、对图像进行二值化处理

图像二值化（ Image Binarization）就是将图像上的像素点的灰度值设置为 0 或 255，也就是将整个图像呈现出明显的黑白效果的过程。

在数字图像处理中，二值图像占有非常重要的地位，图像的二值化使图像中数据量大为减少，从而能凸显出目标的轮廓。

观察 Excel 数字图像的数值变化，找到合适的阈值，对图像进行二值化。

运用到的 Excel 公式：

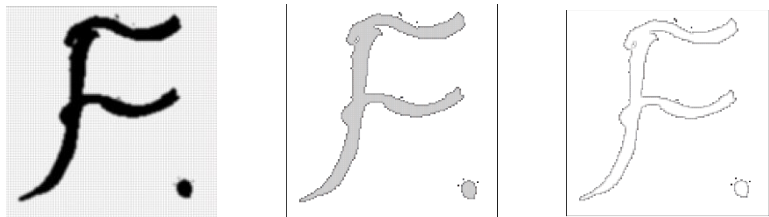
=IF(A1>阈值,255,0)



下图为经过二值化处理以后的图像

五、卷积运算

运用不同的卷积核，可以对图像亮度进行提升，边缘加强或边缘柔化等操作。尝试在 Excel 中运用卷积运算，对图像进行进一步处理。



经过不同卷积核处理以后的图像

六、池化运算

由于现代手机或者相机拍摄的图像分辨率较高，所需存储空间较大。对于作业上传平台来说，不需要如此高的分辨率。我们将采取一定手段进行降采样。在本章节中，我们所用的方式为池化。

池化与卷积不同，由于卷积运算所得到的像素除了最外层一个单位舍弃的像素以外，其余像素与原图一一对应。而池化运算在每 2×2 个像素单元中选取一个像素就形成了 3 个被舍弃的像素。画面上会有空白单元格出现。

思考：如何在 Excel 中去除这些单元格？

用到 Excel 知识点：

IF 函数 筛选 自动填充柄 转置粘贴

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
40																										
41																										
42																										
43																										
44																										
45																										
46																										
47																										
48																										
49																										
50																										
51																										
52																										
53																										
54																										
55																										
56																										
57																										
58																										

部分需要删除的行与列

本章小结

选取一个你熟悉的 App，查看一下它的哪些功能用到了我们本章节学习的算法？

App/算法	二值化	卷积算法	池化算法

线性回归

课前导学

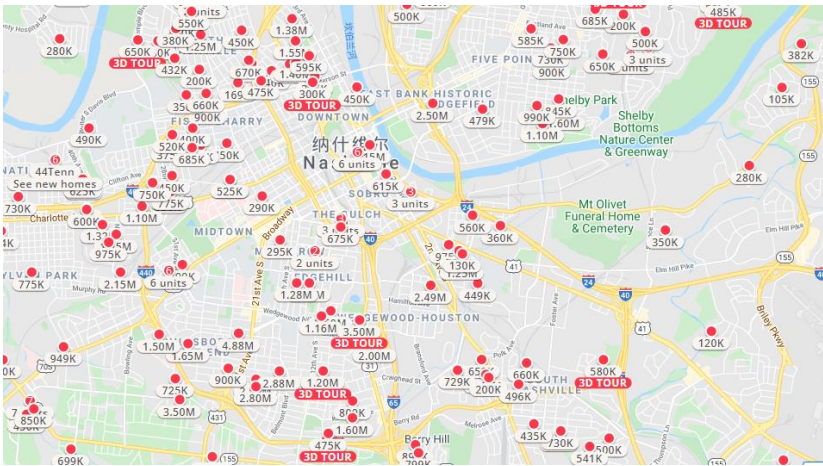
如果地理位置相同有两个房子，一个 50 平方米一个 100 平方米，猜猜看哪个总价更高？你想必会说 100 平方米的房子更贵，这就是人类在生活中形成的一个认识——“房子越大价格越高”。但人的认识总是模糊的，那么现在我们把这个规则教给电脑，同时把该地区所有房子的信息和价格全部都告诉电脑，希望电脑能够根据房子的大小预测房子的价格。在计算机世界，我们把这样的规则叫做回归（Regression）。

学习目标

- 1. 了解麦卡洛克-皮茨模型，理解感知器函数
- 2. 学习 Excel 图表制作方式
- 3. 用一元一次函数模拟纳什维尔的房价
- 4. 学习 GeoGebra 拟合线性回归曲线
- 5. 理解罗森布拉特感知器拟合过程

项目学习指引

在美国的田纳西首府纳什维尔的市中心，挂牌交易的二手房屋的价格与面积如下左图 所示。现在该地区的某位房东手中有一套 1900 平方英尺的房屋，他很想知道能挂牌多少价格，请同学们帮他一起来预测一下。他的房子在当下的市场能卖多少钱？



	面积(sqft)	价格(usd)
1	865	\$615,000
2	670	\$499,900
3	488	\$299,900
4	1347	\$1099,000
5	2903	\$2,150,000
6	2250	\$1,800,000
7	1190	\$955,000

一、数据的采集与整理

在地图上的一堆数据中，我们很难找出规律。所以，在详细询问了该房屋的位置以后，我们选取了在该房屋周围的几所装修、朝向、采光和结构都大致相似的几所住宅进行了价格与房屋面积的整理。我们找到了其中 7 所与该房屋比较近似的住宅价格，整理成数据表，如右图显示。显然，有了这样一张表单以后，可以使查找的范围缩小很多，但似乎还是很难从表中直接看出价格与面积之间的关系。

思考：

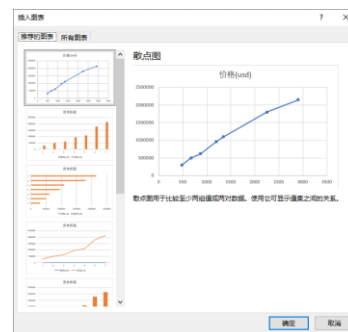
在 Excel 中将数据输入表单，并完成数据的排序。

二、制作数据图表

为了更直观地观察数据之间的关系，我们使用 Excel 制作图表。

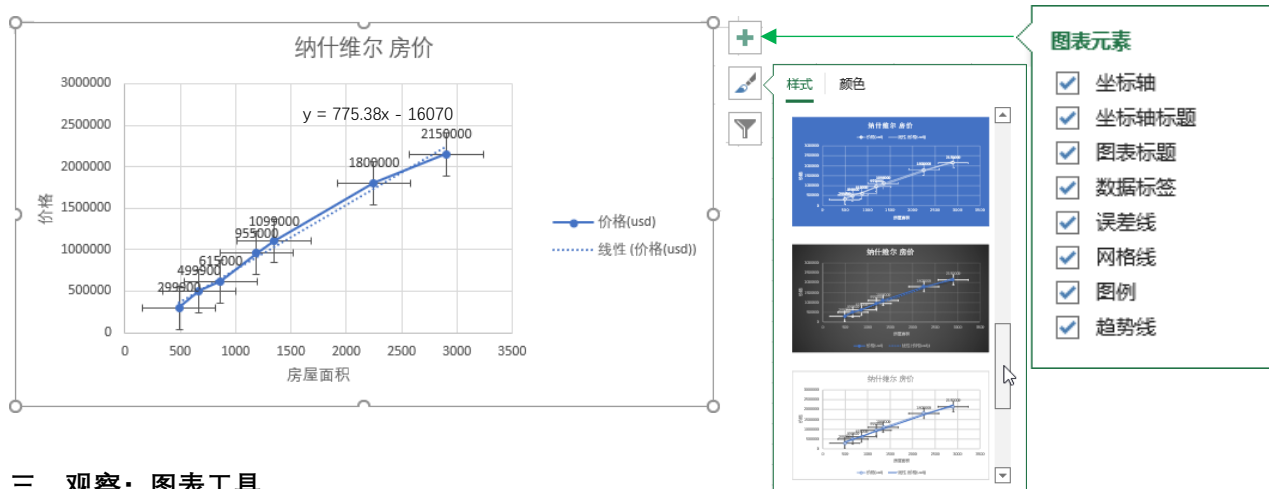
1、选中数据区域，点击“插入”标签→“推荐的图表”进入制作环节。

面积(sqft)	价格(usd)
488	299900
670	499900
865	615000
1190	955000
1347	1099000
2250	1800000
2903	2150000



2、对图表进行进一步设置

利用丰富的设置项，对图表进行美化/设置。



三、观察：图表工具

查看图表工具→图表设计标签，尝试里面的每一个功能并记录你认为重要的：



添加图表：

快速布局：

更改颜色：

数据：

图表样式：

更改图表类型：

移动图表：

四、价格预测

至此，我们已经绘制了价格与面积的关系曲线，但是我们依然无法准确的知道该房东的房屋究竟挂牌价格是多少。要计算出这一价格，还需要更为专业的数学软件。

1、使用或者下载软件

使用在线平台：

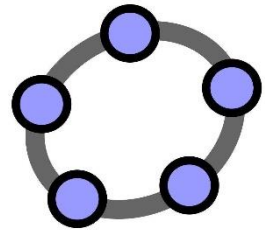
<https://www.geogebra.org/classic>

或者下载 **GeoGebra 经典 6** 应用

官方：<https://www.geogebra.org/download>

学校：<http://ai.sfls.cn/software>

适用于 iOS、Android、Windows、Mac、Chromebook 和 Linux 等操作系统的离线 GeoGebra 应用

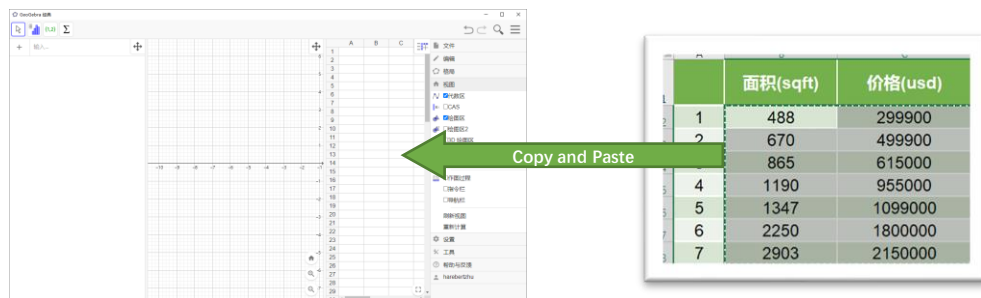


GeoGebra 是一个结合「几何」、「代数」与「微积分」的动态数学软件，它是由美国佛罗里达州亚特兰大的数学教授 Markus Hohenwarter 所设计的。



2、绘制回归曲线预测价格

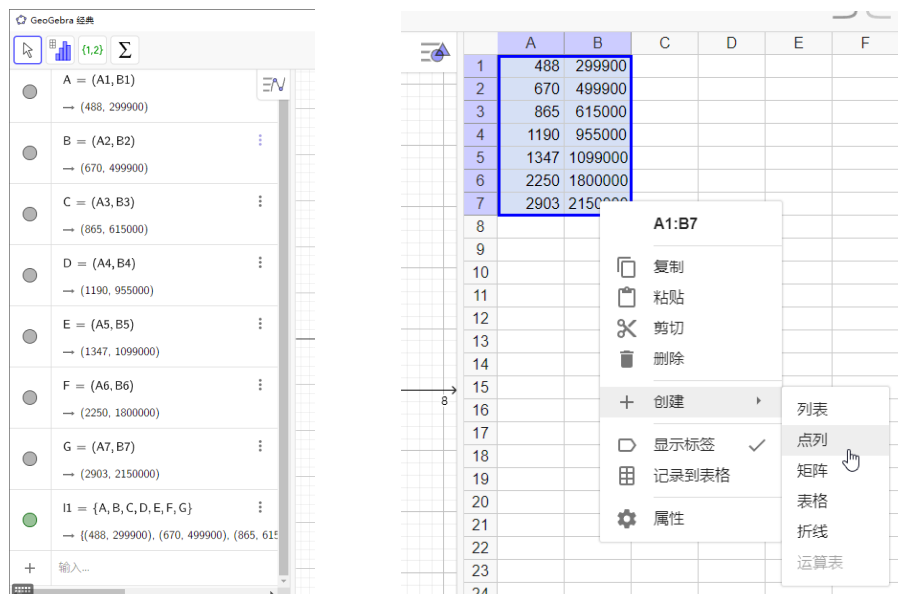
- 1) 打开 **GeoGebra 经典 6** 在菜单的视图标签中打开表格区选项并将 Excel 中的数据粘贴到软件表格中。



- 2) 将复制得到的数据转换成二维点阵列。

选中所有数据，右键选择“+创建”→“点列”；

可以看到在左边的列表中，出现了 A~G 等 6 个点的坐标和一个点列：l1



思考：

为什么在二维图形区域中，看不到绘制的点？有什么办法？

3) 制作拟合函数图像

语句 1:

Fit(<List of Points>, <List of Functions>)

Calculates a linear combination of the functions that best fit the points in the list.

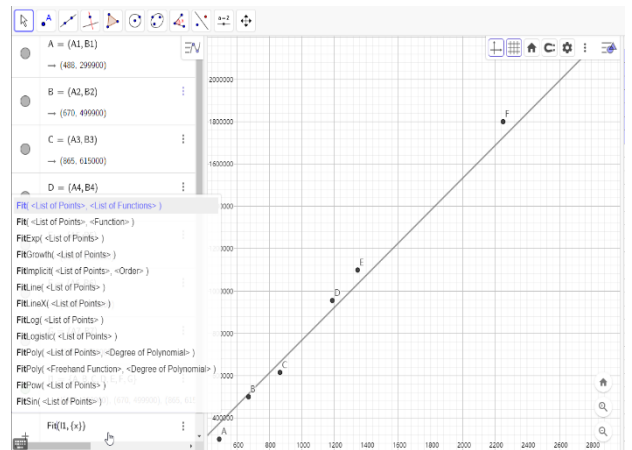
$$\text{Fit} (l1 , \{ x \})$$
$$y = 551.93 x$$

语句 2:

FitLine(<List of Points>)

Calculates the y on x regression line of the points.

$$\text{FitLine} (l1)$$
$$y = 208.77x + 720769.24$$



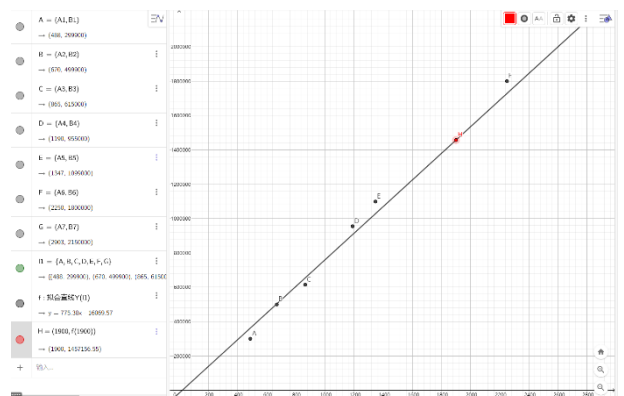
4) 将房屋面积代入函数进行预测并绘制点：

语句：

Point(<List of Points>)

将指定 <串列> 转为一点。

$$\text{Point} (\{ 1900 , f(1900) \})$$



思考：

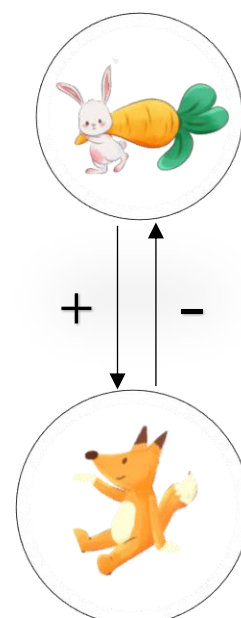
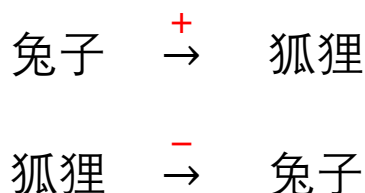
除了预测房价以外，线性回归还能用在哪些地方？

五、线性回归

1. 相关性分析

在辽阔的大草原上，生存者狐狸和兔子，狐狸依赖兔子为生，当兔子数量增长的时候，狐狸的数量也增长，当兔子数量下降的时候，狐狸的生存就会遇到一些困难，因此数量也会下降；反之，当狐狸数量增长的时候，兔子的数量就会减少，当狐狸数量减少的时候，兔子的数量就会增加。

我们可以将狐狸和兔子的数量作为一对相关变量来进行分析：



相关性分析是指对两个或多个具备相关性的变量元素进行分析，从而衡量两个变量因素的相关密切程度。相关性的元素之间需要存在一定的联系或者概率才可以进行相关性分析。

相关性不等于因果性，也不是简单的个性化，相关性所涵盖的范围和领域几乎覆盖了我们所见到的方方面面，相关性在不同的学科里面的定义也有很大的差异。

思考：

如果上述两个动物换成狼与羊，在某个封闭的环境中，他们的相关性是如何呈现的？

项目活动：设计项目活动系统图谱

如果将兔子作为自变量，狐狸作为因变量，他们是正相关的关系，而反之，他们则呈现负相关的关系。在大自然中，有着许多具有相关性的系统，例如：以大气中二氧化碳浓度为自变量，则会导致气温升高；极地冰层减少等因变量的变化。本项目活动中请同学们以小组为单位，探索一下身边还有哪些互为相关性的变量，并绘制成图谱。

项目学习目标

本项目中，我们尝试利用 LOOPY 在线平台，实现相关性系统图谱的绘制，并以小组为单位，交流分享。完成本项目学习以后，所绘制的相关性系统图谱，需达到以下要求：

1. 相关性系统需要能够持久运作（达到平衡）
2. 在系统图谱上有详细的说明
3. 布局合理，能够运用不同颜色表示出系统的不同层级
4. 系统设计中内容正面积极，能引起观者反思

项目学习指引

一、分组讨论，头脑风暴

以 3~4 人为一个小组，根据各自感兴趣的话题，进行头脑风暴。由组长负责，汇总各人的想法，并形成大致草图。可供选择的课题有：

1. 草原上食物链之间的关系

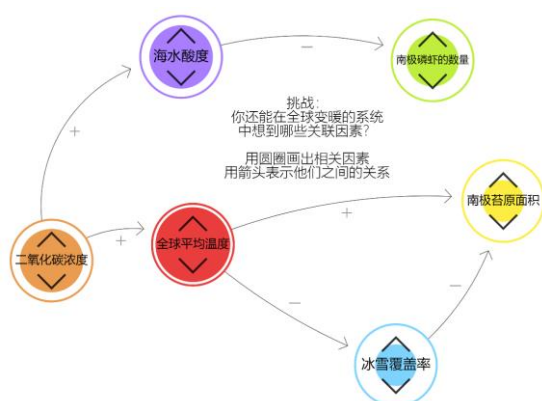
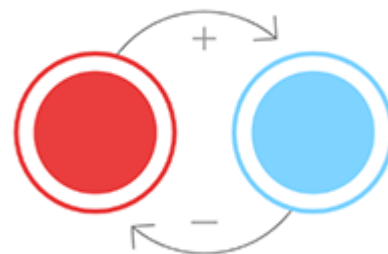
2. 各国的经济实力及相互影响
3. 大气中湿度与温度变化的影响与关系
4. 物理学或化学中各因素/元素相互之间的影响关系

注意：选题无需过大，影响因素不易过多。确定了相互关系之后，可以先绘制草图尝试运行，然后逐渐增加影响因素。

二、登录网站，绘制图谱

登录网址：<http://ai.sfls.cn> 选择相应模块，进行相关性系统图谱设计与制作。在制作过程中，需要注意：

1. 可以参照实例中的图谱进行改制。
2. 图谱能清晰反映各因素之间相互影响的关系，不宜过密。
3. 在图谱中使用 label 简要说明所研究的问题。
4. 相应模块尽量简化，不要追求全面。
5. 完成图谱后进行填写班级、组员名字，进行递交。每个小组可以多次递交，以其中一个作品进行评分。



完成图谱后思考：

小组所绘制的相关性系统图谱是否能够用模型的方式解释整个系统的运作。

在系统中各因素相关运作最后能否达到平衡。

三、分组交流，投票评分

以小组为单位进行交流，交流主要介绍本小组所制作的相关性系统图谱，并运行。

交流结束以后，每位同学（包括小组成员）投票对该作品进行打分，打分一共分为 A~E 五档，其中

- A 档 - 4 分
- B 档 - 3 分
- C 档 - 2 分
- D 档 - 1 分
- E 档 - 0 分

积分规则为：(n 为总投票人数， $n_A \sim n_E$ 分别为每档人数)

$$Score = \frac{100}{4n} \times (4 \cdot n_A + 3 \cdot n_B + 2 \cdot n_C + n_D)$$

2. 回归分析

问题提出：

在上一课的相关性分析中，同学们发现虽然自变量与因变量呈现正相关或者负相关的关系变化，但是他们的变化速度是一致的。而在自然界中，实际情况并非如此，当一对相关变量 A 和 B 呈正相关或者负相关时，他们的增长率并不一定是相同的。例如在我们之前“纳什维尔房价问题”中，房价（美元）和面积（平方英尺）这一对相关变量的增长率并不一致，平均每 1 平方英尺的面积增长会导致约 800 美元左右的房价增长。这个时候，我们需要考虑不同增长率之间的比值，在本节中，我们用 k 来表示。

如果我们将自变量定义为 x ，因变量定义为 y ，一个相关性相对稳定的系统会呈现以下关系式：

$$y = kx$$

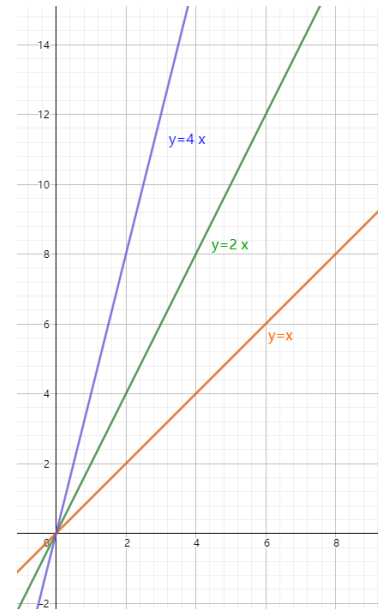
我们可以理解为对于每一个自变量 (x)，都会对应一个因变量 (y)，而自变量 (x) 与因变量 (y) 的关系可以用比例系数 (k) 来确定。
在 GeoGebra 上用一条直线可以表示这样的关系式。

在图中，不同的 k 影响到不同的直线倾斜程度（斜率）例如

当 $k=4$ 时 - 蓝色线条

当 $k=2$ 时 - 绿色线条

当 $k=1$ 时 - 红色线条



思考：

当有一组因变量 x ，使用哪一条直线（选取哪一个 k 值）来预测这一组自变量所对应的因变量呢？

假设：

有一对已知的关系变量， $x=2$ ； $y=4$ ，对这一组变量采用以上三条直线进行预测，结果如下：

当 $k=4$ 时 - y 的预测值为 8

当 $k=2$ 时 - y 的预测值为 4

当 $k=1$ 时 - y 的预测值为 2

很显然，在这次预测中， $k=2$ 似乎与实际因变量 y 值最为接近。

在实际操作中，不一定每一个点都能“拟合”到直线上，我们必须考虑当每一个 k 值所造成的误差。

这个误差，在数学上，对于每一对 x 和 y 的实际值来说，其误差可以这样来表示：

$$(d_{\text{误差}})^2 = (y_{\text{实际}} - y_{\text{预测}})^2 = (y_{\text{实际}} - k \cdot x_{\text{实际}})^2$$

注：

其中，误差 (d) 表现为二维坐标图中点到直线 $f(x)$ 的垂直距离

用误差的平方值，可以消除正负值对误差的影响，只保留绝对误差。

这样，对于每一个不同的 k 值，我们都可以用数据来直观地描述/观察它对预测值的影响。

动手尝试：

登录网址：<http://ai.sfls.cn> 选择相应模块进行 k 值测试

在默认情况下，直线 $y=kx$ 中，k 值取 1.

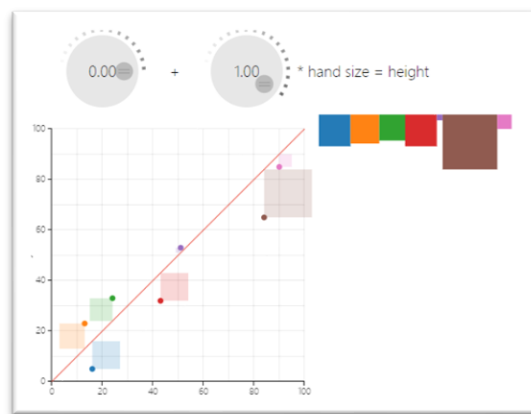
通过拖动灰色圆盘，调整 k 值。

在二维坐标系中，每一个实际点的误差平方表现为：

以该点到直线的垂直距离为边长的正方形的面积。

最优的拟合方案为：

通过调整 k 的值，使所有正方形的面积之和（误差平方和）为最小



通过观察，完成以下问题：

通过拖动 k 值灰度圆盘（将 k 值增大/减小），直线会如何进行变化？

当 k 值增大时：

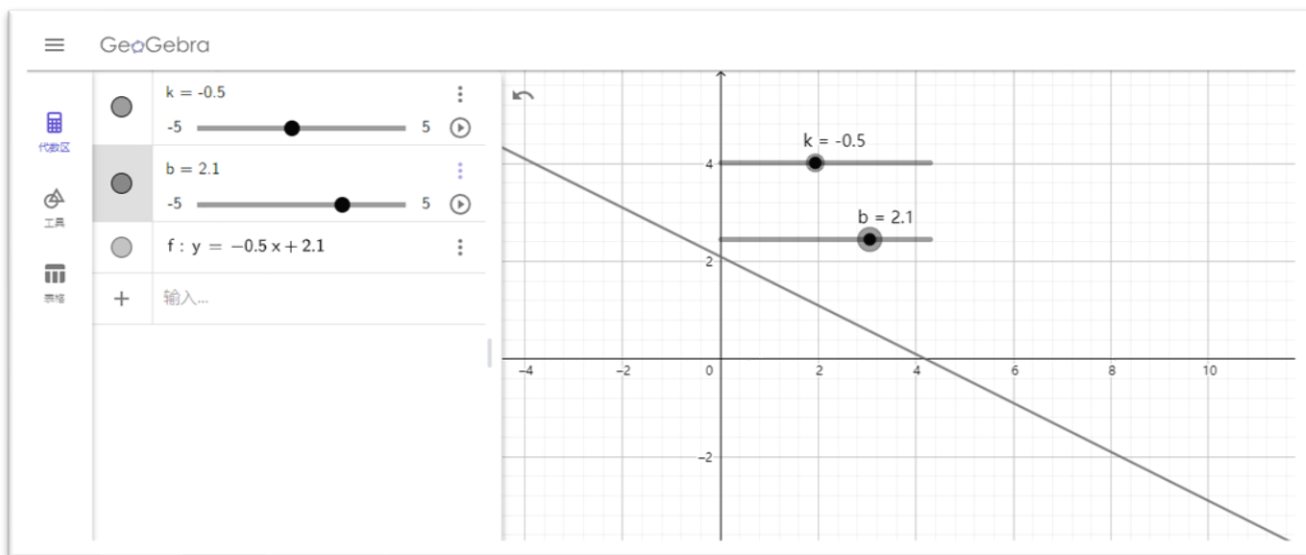
当 k 值减小时：

除了 k 值以外，还有一个灰度圆盘，我们记做 b 值，当拖动这个圆盘时，直线会如何进行变化？

当 b 值增大时：

当 b 值减小时：

在 GeoGebra 中输入 $y = kx + b$ 验证你的结论。



回归分析中，只包括一个自变量和一个因变量，且二者的关系可用一条直线近似表示，这种回归分析称为一元线性回归分析。如果回归分析中包括两个或两个以上的自变量，且因变量和自变量之间是线性关系，则称为多元线性回归分析。

六、麦卡洛克-皮茨模型

在疫情期间，每次做核酸检测的时候，鼻腔/咽喉中的神经末梢会向大脑传递信号：有异物入侵了。然后大脑会做出反应，例如咳嗽、流泪。

对于人来说感知器就是我们大脑中的一系列神经元，把他们抽象成一个神经元，由输入端、神经中枢等组成。1943 年美国心理学家麦卡洛克 (McCulloch, W. S.) 和数学家皮茨 (Pitts, W.) 等提出利用神经网络对信息进行处理的数学模型，从此人们开始了对神经网络的研究。该模型的主要目的是完成对神经元状态的描述。

这个模型利用数学函数中的自变量和因变量模拟生物神经网络中的树突和轴突，而不同的权值 ω 控制着每个输入对输出的影响

如果设输入信号为 x_i ，权值为 ω_i ，输出为 y

如果有三个输入信号，则可以列出这样的关系式：

$$y = \omega_1 x_1 + \omega_2 x_2 + \omega_3 x_3$$

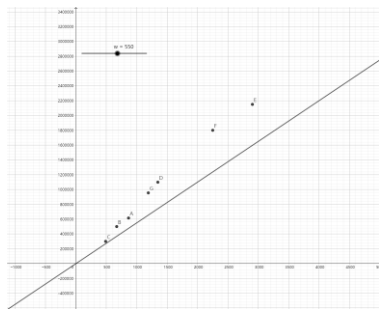
如果有多个输入信号，可以将关系式写成如下形式：

$$y = \sum_{i=1}^n \omega_i x_i$$

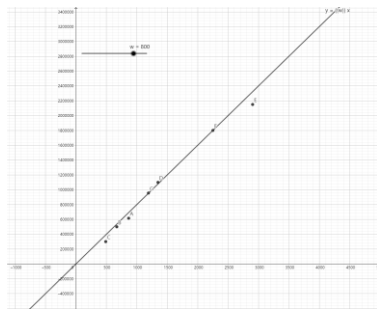
对于纳什维尔的房价问题，我们运用麦卡洛克-皮茨模型，将房价作为输出信号、面积作为输入信号，可以得到这样一个关系等式：

$$y = \omega x$$

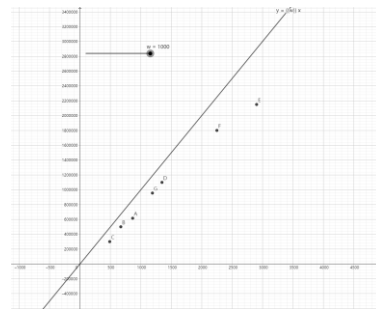
当 ω_1 处于不同值的时候，在 **GeoGebra 经典 6** 上做出的图像如下所示：



$\omega = 550$



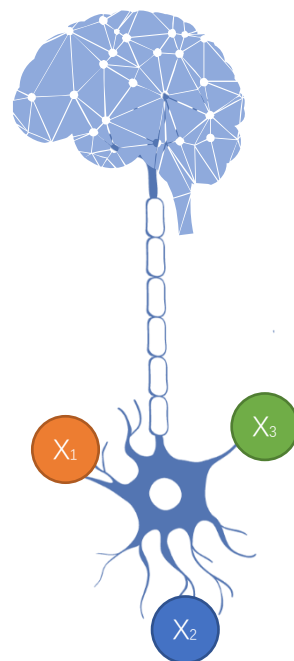
$\omega = 800$



$\omega = 1000$

思考：

如果随机给定一个 ω 的值，在绘制出直线以后，如何根据预测值和实际值的误差调整 ω 的值？



我们可以将 ω 看做**预测房屋单价**，那么在不同单价 ω 的预测下，每一个房屋的实际价格与预测价格的差值为：（以 A 点(865,615000) 为例，当 $\omega = 550$ 时）

$$\begin{aligned}\Delta y_1 &= y_1 - \omega x_1 \\ \Delta y_1 &= 615000 - 550 \times 865 \\ \Delta y_1 &= 139250\end{aligned}$$

根据每一个 Δy_n 来调整 ω 的值，我们有以下公式：

$$\omega_{new} = \omega_{old} + \alpha \cdot \Delta y_i \cdot x_i$$

其中 α 称之为 学习率，通常 $0 \leq \alpha \leq 1$

思考 1:

当我们利用以上公式对每一对 x_i 和 y_i 进行计算，每一次对 ω 的值进行更新。经过一轮迭代（Iteration）以后， ω 的值已经比较接近在上一课中使用 **GeoGebra 经典 6** 的拟合功能求得的 ω 值了。

右侧表格中，模拟的是当 $\alpha = 10^{-7}$ 时一轮迭代的学习情况，通过 Excel 软件，模拟不同的 α 值。并选取最优的 α 值，说明原因。



x	y	ω	Δy
865	615000	550	139250
670	499900	562.0451	123329.8
488	299900	570.3082	21589.59
1347	1099000	571.3618	329375.7
2903	2150000	615.7287	362539.6
2250	1800000	720.9739	177808.6
1190	955000	760.9809	49432.75
		766.8634	

当 $\alpha = 10^{-7}$ 时的学习情况

你选择的 α 值是_____

原因是：_____

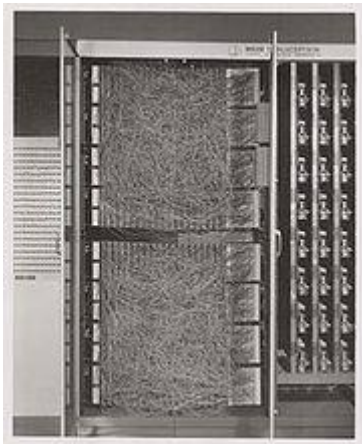
思考 2:

除了修改 α 学习率以外，还能通过什么方式达到比较理想的 ω 值？

知识拓展:

1943 年，心理学家沃伦·麦卡洛克和数理逻辑学家沃尔特·皮茨在合作的《A logical calculus of the ideas immanent in nervous activity》论文中提出并给出了人工神经网络的概念及人工神经元的数学模型，从而开创了人工神经网络研究的时代。1949 年，心理学家唐纳德·赫布在《The Organization of Behavior》论文中描述了神经元学习法则——赫布型学习。

人工神经网络更进一步被美国神经学家弗兰克·罗森布拉特所发展。他提出了可以模拟人类感知能力的机器，并称之为‘感知机’。1957 年，在 Cornell 航空实验室中，他成功在 IBM 704 机上完成了感知机的仿真。两年后，他又成功实现了能够识别一些英文字母、基于感知机的神经计算机——Mark1（右图），并于 1960 年 6 月 23 日，展示与众。



基于感知机的神经计算机——Mark1